CAN4Age: Chinese affective norms for 4-character words rated by older and younger adults

Pingping Liu[1,2], Qin Lu[3], Zhen Zhang[1,2], and Buxin Han[1,2]

[1] Center on Aging Psychology, CAS Key Laboratory of Mental Health, Institute of Psychology, Beijing, China

[2] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

[3] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Author Note

This research was partially supported by the National Natural Science Foundation of China (No. 31600887), the Hong Kong Scholars Program (No. XJ2015050), and by the CAS Key Laboratory of Mental Health (No. KLMH2014ZG14).

Address of correspondence:

Pingping Liu

Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing, People's Republic of China

Email: liupp@psych.ac.cn

Reading head: Age differences in affective norms for Chinese words

Abstract

Age-related differences in affective meanings of words are widely used by researchers studying emotions, word recognition, attention, memory and text-based sentiment analysis. However, no Chinese affective norms for older adults are available. This article firstly presents the available large-scale Chinese affective norms for 2, 061 4-character words rated in labs by 114 older and 150 younger adults (CAN4Age) who evaluated these words on four dimensions: valence, arousal, dominance, and familiarity. We also compiled 4 lexical variables for each word, including word frequency, word complexity, character frequency and character complexity. In general, older adults tend to evaluate emotional words more extremely than younger adults do. That is, they rate positive words as more positive and negative words as more negative than younger adults do. Specifically, older adults tend to perceive positive words as more arousing and less controllable and negative words as less arousing and more controllable than that of younger adults. This age-related database will enable researchers to study how emotional characteristics of words influence their cognitive processing, and how this influence evolves with age in Chinese. This age-related difference study on affective norms not only provides insights to cognitive neuroscience, gerontology and psychology in experimental studies, but the produced affective word collection also has great value as a resource for affective analysis in natural language processing applications. These norms can be downloaded as supplemental materials with this published article.

Keywords: Affective norms, ratings, age differences, Chinese

Introduction

Affective ratings of words are in high demand because they are widely used by researchers studying emotions and moods (Wolf & Demiray, 2019), word recognition (Citron, Weekes, & Ferstl, 2013; Kuchinke & Mueller, 2019; Kuperman, Estes, Brysbaert, & Warriner, 2014), memory (Garrison & Schmeichel, 2019; Majerus & D'Argembeau, 2011; Monnier & Syssau, 2008), attention (Mathewson, Arnell, & Mansfield, 2008), and text-based sentiment analysis (Kratzwald, Ilić, Kraus, Feuerriegel, & Prendinger, 2018; Warriner, Kuperman, & Brysbaert, 2013). In recent years, the role of age in modulating the processing of emotional information has become a focus of increasing interest in the field of life span psychology (English & Carstensen, 2014a; Mather & Carstensen, 2005; Notthoff & Carstensen, 2014; Reed, Chan, & Mikels, 2014; Steenhaut, Demeyer, De Raedt, & Rossi, 2018; Stine-Morrow, Miller, & Hertzog, 2006; Wirth, Isaacowitz, & Kunzmann, 2017). Because of life experience and age-related biological changes, differences in the perception of words by age may be ubiquitous in terms of affective polarity, arousal, and control. However, such age-normative information remains scarce. Little is known about age-related differences in the perception and meaning of emotional words (Fairfield, Ambrosini, Mammarella, & Montefinese, 2017; Gilet, Grühn, Studer, & Labouvie-Vief, 2012; Grühn & Scheibe, 2008; Grühn & Smith, 2008; Ready, Santorelli, & Mather, 2017). The aim of this study was to close this gap for Chinese and provide an age-adapted tool for future research on the processing of emotion words from a developmental point of view. A new affective lexicon as a database (i.e., Chinese affective norms for

4-character words rated by older and younger adults, CAN4Age) is obtained in this work. Its construction is based on Bradley and Lang's (1999) procedure and the lexicon contains emotional ratings and familiarity by older and younger adults for 2,061 words.

Most norming studies on emotions and languages have been based on Bradley and Lang's (1999) Affective Norms for English Words (ANEW) database (Moors et al., 2013; Warriner et al., 2013). Three types of ratings were carried out for 1,034 English words in this database, which was developed within the dimensional theory of emotions (Osgood, Suci, & Tannenbaum, 1957; Russell, 2003; Wundt, 1912/1924). The first dimension of ratings concerns the valence (or pleasantness) of emotions elicited by a word (going from unhappy to happy). The second dimension measures the degree of arousal which reflects the subjective level of activation or intensity that a word evokes (ranging from calm/quiet to excited/active). The third dimension is dominance which refers to the degree of control exerted by a word (ranging from weak/submissive to strong/dominant).

Age differences in emotional functions

Could valence, arousal and dominance ratings of younger adults be generalized to older age group? Many studies on aging and emotions have indicated that older and younger adults differ in several aspects of emotional functions (English & Carstensen, 2014a; Mather & Carstensen, 2005; Notthoff & Carstensen, 2014; Reed et al., 2014; Steenhaut et al., 2018; Wirth et al., 2017). First, emotional experience appears to grow

more positive with age (i.e., positivity effect). Older adults attend less to negative information and more to positive information compared to younger adults based on cross-sectional and longitudinal studies (Carstensen, 2006; English & Carstensen, 2014a, 2014b; Kunzmann, Little, & Smith, 2000; Mather & Carstensen, 2005; Mroczek & Kolarz, 1998). Second, older adults tend to report having better developed emotion regulation abilities than younger adults do. They appear to dissipate negative affect more effectively, and focus more on self-control of their inner emotions than younger adults do (English & Carstensen, 2014b; Grühn & Scheibe, 2008; Hess, Popham, Dennis, & Emery, 2013). Third, there is some evidence that older adults tend to show reduced autonomic reactions to emotional stimuli compared to younger adults (Ferrari, Bruno, Chattat, & Codispoti, 2017; Keil & Freund, 2009; Steenhaut et al., 2018; Streubel & Kunzmann, 2011; Uchino, Birmingham, & Berg, 2010). Overall, a recent meta-analysis by Reed and colleagues (2014) has confirmed that older adults tend to show a significant bias toward positive versus negative information, whereas younger adults show the opposite pattern. Thus, these age-related differences in emotional experience, control, and reactivity suggest that emotional ratings of younger adults could not be generalized to older adults.

Empirical evidence for age-related differences in subjective evaluations of emotional words remains scarce and available mainly for pictorial material. Some previous studies have obtained emotional ratings of standardized pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1998) between older and younger adults, with inconsistent results. There were reports of

older adults showing lower subjective ratings of their feelings than younger adults (Keil & Freund, 2009; Streubel & Kunzmann, 2011). Yet, in other experiments, there were opposite results (Gavazzeni, Wiens, & Fischer, 2008; Grühn & Scheibe, 2008; Grühn & Smith, 2008; Steenhaut et al., 2018) or similar ratings (Ferrari et al., 2017; Wieser, Muhlberger, Kenntner-Mabiala, & Pauli, 2006). Possible mechanisms under these age related inconsistencies have yet to be well-established (Steenhaut et al., 2018). Furthermore, some studies have revealed neural processing of emotional pictures and words are different (Kensinger & Schacter, 2006; Leclerc & Kensinger, 2011). Thus, measurements of other types of age-related emotional stimuli, especially words, would bring some clarity to age-related differences in emotional reactivity.

A few studies have examined changes of self-reported affective responses to standardized words between older and younger adults in German (Grühn & Smith, 2008; Keil & Freund, 2009), French (Gilet et al., 2012), English (Ready et al., 2017), and Italian (Fairfield et al., 2017), and they showed substantial age differences. Grühn and Smith (2008) supplied the AGE database (Age-dependent evaluations of German adjectives) of 200 words which were evaluated by older and younger adults. A large proportion of words show age-related differences in valence (30% of all 200 words), arousal (21%), and dominance ratings (16%). In general, older adults rate positive words as more positive, more arousing and less controllable than younger adults do. In contrast, they tend to rate negative words as less arousing and more controllable than younger adults (Grühn & Smith, 2008). However, Ready and colleagues (2017) observed that older adults tend to rate negative words as more activating than younger

adults based on a sample of 70 emotion terms. Gilet and colleagues (2012) also demonstrated that older adults tend to rate negative words as more arousing than younger adults do. They reported a stronger association between valence and arousal ratings with age such that negative valence is more strongly associated with high arousal for older adults. Thus, age-related differences in emotional meanings of words (e.g., negative words) seem to vary in different language and culture environment.

Affective ratings and languages

It is still an open question whether age-related differences in affective ratings would be similar in different languages or cultures. Norms of affective properties of words are available in a number of languages, such as English (Eilola & Havelka, 2010; Stadthagen-González & Davis, 2006; Stevenson, Mikels, & James, 2007; Warriner et al., 2013), French (Gilet et al., 2012; Monnier & Syssau, 2017), German (Grühn & Smith, 2008; Kanske & Kotz, 2011; Schmidtke, Schröder, Jacobs, & Conrad, 2014), Spanish (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017; Ferré, Guasch, Moldovan, & Sánchez-Casas, 2012; Guasch, Ferré, & Fraga, 2016; Hinojosa et al., 2016; Stadthagen-Gonzalez, Imbault, Sánchez, & Brysbaert, 2017), Portuguese (Soares, Comesaña, Pinheiro, Simões, & Frade, 2012), Dutch (Moors et al., 2013), Italian (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014), and Chinese (Ho et al., 2015; Liu, Li, Lu, & Han, 2018; Y. Wang, Zhou, & Luo, 2008; Yao, Wu, Zhang, & Wang, 2017). It is widely known that there are some differences between western and eastern cultures in areas including age-related personality, social relationships, and cognition (Fung, 2013; Markus & Kitayama, 1991). Although

age-related differences in emotional ratings have been demonstrated in western countries (e.g., France, Germany, and USA), these detailed findings may not be generalizable to Chinese.

Limitations of conventional Chinese affective norms

To our best knowledge, no Chinese affective norms for older adults are available. Although a large proportion of the world's population consists of older adults (more than 240 million) in China, little is known about age-related differences in subjective evaluations of emotional words in Chinese. For instance, Y. Wang and colleagues (2008) provides the valence, arousal and dominance ratings for 1,500 2-character words, and Yao and colleagues (2017) collected valence and arousal ratings for a total of 1,100 2-character words by a paper-and-pencil test from university students. According to the *Chinese Lexicon* (2003), 64% of words are 2-character words, and 14% are 4-character words. Few previous Chinese affective lexicons are based on 4-character words, which convey more complex and abundant meanings than 2-character words. Recently, Liu and colleagues (2018) described an annotated dataset on valence and arousal for a large lexicon of 2,076 4-character Chinese words rated by younger adults. However, the database of Liu et al. (2018) did not provide ratings on all three dimensions listed in ANEW, that is, the dominance dimension was not rated. Dominance/power has not often been included in previous word norming studies (but see Bradley & Lang, 1999; Grühn & Smith, 2008; Moors et al., 2013; Warriner et al., 2013), even though it has been identified as an important variable in emotion studies in addition to valence and arousal (e.g., Fontaine et al., 2007; Osgood

et al., 1957). In a word, extant databases do not include ratings for older adults, and the lexicon for 4-character words remains scarce in Chinese.

The present study

In order to address the research gap on age-related differences in evaluations of emotional words in Chinese, the present study obtained affective norms with well-designed procedures in labs for 2,061 4-character Chinese words from both older and younger adults. Making these abundant age-related affective databases available could help improve the performance of emotional word recognition models. Currently, models of word recognition have not be integrated with affective features (Citron, Weekes, & Ferstl, 2014; Kuperman et al., 2014). Furthermore, the collection of 4-character words can serve as a supplemental resource to the currently available 2-character affective lexicon for Chinese sentiment computing. Our database also provides raw data to enable researchers to study how emotion influences cognitive processing, and how this influence evolves with age.

Method

Participants

One hundred and twenty-five older adults (56-85 years of age, 50.4% female) and 160 younger adults (16-40 years of age, 50% female) from the local community or campus were recruited through advertisements in Beijing for this study. All participants were native Chinese speakers with normal or corrected-to-normal vision, and they received an honorarium of 40 RMB per hour for their participation. The

study was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences. Some of the younger cohort is the same as those described in Liu et al. (2018). However, this study recruited additional younger adults and the new older cohort in order to collect sufficient data to study age differences. In order to screen for possible mild cognitive impairment, all participants need to complete a battery of neuropsychological tests. The participants' demographic details and their self-rated health information were collected firstly. Then all participants were given the Mini Mental State Examination (i.e., MMSE) as a preliminary screening measure, and the minimum score of 26/30 was required (Folstein, Folstein, & McHugh, 1975). The battery comprised of the Digit Span Forward and Digit Span Backward (Wechsler, 1981), the Vocabulary Test (Wechsler, 1981), and the Category Fluency Test (Spreen & Strauss, 1998). These tests were used as indications that the participants had intact cognitive abilities, and they were administered in a separate 0.5 hour session. Seven older adults and one younger adult were removed because of their low education or lower scores in neuropsychological tests (i.e., MMSE and Vocabulary Test). One older adult and 9 younger adults were removed because of a high number of outlier scores or because the ratings given by them seem to be quickly entered at random. Three older adults were removed because they could not use a computer to complete the experimental task. The final sample consists of 114 older adults (56-85 years of age, M = 70.05, SD = 6.01; 54% female) and 150 younger (16-38 years of age, M = 21.59, SD = 3.41; 50% female), and they were free from

neurological and psychiatric disorders. The demographic characteristics of the final 264 participants are presented in Table 1.

Insert Table 1 here

Materials

A collection of 2,290 4-character words from the work of Liu et al. (2018) provided the raw data for this study. This collection includes the affective norms from younger adults of a large scale with valence and arousal ratings. However, only 2,061 words are in the final set in this work. 15 words were removed due to typographical errors and 214 words were removed as they are marked as "not known" by more than 10% of all participants in that study. According to the *Chinese Lexicon* (2003), the mean word frequency of the final 2,061 words was 135 (*SD* = 259, range = 2 to 5,384, median = 69) occurrences per million, and the mean word complexity of the set was 30.63 (*SD* = 7.42, range = 8 to 72, median = 30). The average frequencies of the first, second, third and fourth characters were 1,076 (*SD* = 1,902), 1,241 (*SD* = 1,901), 1,277 (*SD* = 1,955) and 1,160 (*SD* = 1,844) occurrences per million, respectively. The average complexity of the first, second, third and fourth characters were 7.76 (*SD* = 3.02), 7.50 (*SD* = 3.08), 7.58 (*SD* = 3.14) and 7.78 (*SD* = 3.08), respectively. These 2,061 words could be considered frequently used, because affective ratings of unfamiliar words are not valid for most participants.

For each word in our final set, each dimension (valence, arousal, dominance, familiarity) was rated from a minimum of 48 (24 older adults) to a maximum of 212 participants (107 older adults). To avoid the interference of the 4 dimensions during data collection, we designed our experiment in two steps. In the first step, a questionnaire (Type 1) was prepared to collect only the ratings of valence and arousal. This is because valence and arousal are more intuitive to people and are easier to rate than the other two dimensions. An option to mark a word as Unknown was also given in Type 1. In the second step, another questionnaire (Type 2) was prepared to collect the ratings of dominance and familiarity (see Fig. 1).

Insert Figure 1 here

In data collection of Type 1 questionnaire, 2,290 words were divided into 6 blocks containing 381 to 382 words in each block for older participants and 5 blocks containing 458 words each for younger participants, given that older adults tend to respond slower than younger adults (Liu, Liu, Han, & Paterson, 2015; Paterson, McGowan, & Jordan, 2013; Rayner, Yang, Schuett, & Slattery, 2014; Shafto & Tyler, 2014; Stine-Morrow et al., 2006; J. Wang et al., 2018). To avoid primacy or recency effects, the order in which words appeared in the block was randomized across participants. Data were then collected from a total of 99 older (56-85 years of age, *M* = 70.52, *SD* = 6.08; 53% female) and 102 younger (16-38 years of age, *M* = 21.85, *SD* = 3.63; 50% female). According to the participants' convenience and demographic

background, thirty-nine older and 41 younger adults completed one block, 36 older and 50 younger adults completed two blocks, 11 older and 8 younger adults completed three blocks, and 13 older and 3 younger adults completed more than three blocks. 214 words were removed based on the collected data in the Unknown option.

In data collection of Type 2 questionnaire, 2,076 words were divided into 6 blocks containing 346 words each for older participants and 5 blocks containing 415 to 416 words each for younger participants. This questionnaire was completed by 46 older (58-81 years of age, *M* = 70.22, *SD* = 5.61; 52% female) and 78 younger participants (17-28 years of age, *M* = 21.09, *SD* = 2.59; 50% female). Five older and 21 younger adults completed one block, 11 older and 33 younger adults completed two blocks, 11 older and 5 younger adults completed three blocks, 9 older and 3 younger adults completed four blocks, and 10 older and 2 younger adults completed more than four blocks.

Procedure

A computer-based questionnaire was used, and participants gave the ratings in labs at the Institute of Psychology of Chinese Academy of Sciences, Beijing, in small age-homogeneous groups of 3-6 persons in the presence of two researchers. After completing informed consent, some demographic questions (i.e., age, gender, education, self-reported health, etc.) and the battery of neuropsychological tests (i.e., MMSE, vocabulary, verbal fluency and digit span) were also collected. Each participant was seated in front of a desktop computer and received an instructions

sheet for the relevant dimensions before starting the rating procedure. At the beginning of each block, the participants were told that they would be presented with a block of words and their task was to rate them along the two dimensions assigned to them (i.e., valence/arousal, or dominance/familiarity, see Fig. 1). All dimensions were rated on 9-point scales. Response scales ranged from extremely unpleasant (1) to extremely pleasant (9) for valence, from extremely calming (1) to extremely exciting (9) for arousal, from extremely controlled (1) to extremely control (9) for dominance, and from extremely unfamiliar (1) to highly familiar (9). They were given instructions with examples and the opportunity to practice 15 trials using the scale to ascertain that participants understood the task. The instructions for the different norms were either adapted on the basis of original instructions taken from previous published studies (Bradley & Lang, 1999; Eilola & Havelka, 2010; Stadthagen-Gonz á lez & Davis, 2006; Stadthagen-Gonzalez et al., 2017; Warriner et al., 2013) or from previous Chinese normative studies (Liu et al., 2018; Y. Wang et al., 2008; Yao et al., 2017). The exact wording in Chinese as well as an English translation are provided in the Appendix.

The paradigm was automated using E-prime (Psychology Software Tools, Inc., Sharpsburg, PA) and stimuli were presented on a computer display. As shown in Figure 1, each trial began with a fixation cross (+) displayed in the center of the screen for 600 ms. Each word was displayed, one at a time along with the respective 9-point scale until participants responded by clicking on the appropriate rating using the computer mouse. Word stimuli were presented on a 17-inch LCD monitor (resolution: 1024 × 768 pixels, refresh rate: 85 Hz) in white on a light gray

background. The contrast was low to minimize eye fatigue. Each 4-character word was displayed on a single line in Courier New 34-point font, and the size of each Chinese character was 84 × 84 pixels. In Type 1, participants rated all of the words first for valence, and then for arousal. In Type 2, participants rated all of the words first for dominance, and then for familiarity. Participants were allowed to stop rating during a rating session and to resume after a short break at their pace. The rating of each block lasted appropriately an hour for older participants and 45 minutes for younger participants. Some participants would complete more than one block, and they were asked to leave at least a 6-hour interval between two blocks. The order of these blocks was counterbalanced across participants.

Results and discussion

1 Data trimming

Altogether, 993,424 ratings and data points of response times (RTs) were collected across all four dimensions. We conducted the following outlier analysis. First, we removed all ratings for words for which at least one participant indicated that the word was unknown to them (3.3% of all). Second, we discarded ratings of participants who gave the same rating for more than 85% of the words for each dimension (0.61% of the collected blocks). Third, we excluded the ratings for 15 words because they were typed incorrectly in the E-prime program (0.069%). Fourth, means and standard deviations (*SD*) were calculated for each word of older and younger participants, respectively. We removed the data for those participants whose

scores were 2.5 standard deviations away from their group's average for each word (3.2% of all). For the final 2,061 words, the data set consisted of 130,960 observations for valence and arousal separately (91% of the original data pool), 100,775 observations for dominance and familiarity separately (96% of the original data pool), 122,187 observations for RTs of valence and arousal rating separately (85% of the original data pool), and 96,860 observations for RTs of dominance and familiarity separately (91% of the original data pool).

2 Availability of CAN4Age database

In the final data set, 99.98% of the 2,061 words had been rated by at least 20 older adults and 20 younger adults for the affective dimensions and familiarity. For each word, we calculated the mean and *SD* for each age group and gender, and compiled the affective ratings and familiarity into a database. The database contains 2,061 entries for the corresponding Chinese words based on Romanized Pinyin order, together with their English translations (based on Google Translate, Baidu Translate and five Chinese-English bilinguals), valence category, rating values, RTs, sample sizes (No. of participants) for valence, arousal, dominance and familiarity. Mean rating values (Mean), mean RT and *SD* of the four dimensions for each word are given for the global sample (All), the older adults, the younger adults, all women and all men, respectively. For each age group, statistical data are also given for females and males separately. The CAN4Age database also contains information about word frequency, word complexity, character frequency and character complexity, which are

taken from *Chinese lexicon* (2003). The full set of norms is available for access in an Excel file as supplementary materials to this published article.

3 Descriptive statistics

Table 2 reports descriptive statistics and group differences for valence, arousal, dominance and familiarity ratings and for each age group and gender. Figure 2 shows the distributions of the ratings of the four dimensions for older and younger adults. The distributions of valence, dominance, and familiarity ratings are negatively skewed for both older (G_1 = -.20, -.069, and -.69, respectively) and younger adults (G_1 = -.092, -.14, and -1.29, respectively). On the other hand, arousal is positively skewed for older (G_1 = .60), but negatively skewed for younger adults (G_1 = -.24). 51% and 49% of 2,061 words are rated above the middle of the valence rating scale (i.e., 5) for older and younger adults, respectively (no significant age effects: χ^2 = 1.71, p = .19). 84% and 72% of the words are rated above the middle of the arousal rating scale for older and younger adults, respectively (significant age effects: χ^2 = 82.63, p < .001). Older adults' arousal responses are distributed in a smaller range (4.5-7.0) than of the younger adults (3.5-7.5). 36% and 52% of the words are rated above the middle of the dominance rating scale for older and younger adults, respectively (significant age effects: χ^2 = 104, p < .001). It indicates that younger adults are more likely to judge words in control, as compared to older adults.

Insert Table 2 and Figure 2 here

Figure 3 shows plots of the means and *SDs* of the ratings for all dependent variables for older and younger adults. Ratings of valence are relatively stable across participants, while arousal, dominance and familiarity are much more divergent (see *AvgSD* in Table 2). This is also indicated by the difference between the average standard deviations of the dimensions for the global sample: 1.13 for valence, 1.48 for arousal, 1.77 for dominance and 1.33 for familiarity, respectively. Similar to the patterns reported by Moors et al. (2013), the scatterplot for valence (Fig. 3a-b) shows that there are two types of words in the midrange (around the score of 5.0): (a) words with low *SDs* upon which participants agree that they are neutral such as the word for *Tropic Capricorn* (南回归线 / Nan2hui2gui1xian4), and (b) words with high *SDs* that evoked both high and low values from different participants. For instance, this word *慷慨就义* (Kang1kai3jiu4yi4) for *Go to one's death like a hero* is rated negative by 37% of all participants contracting 41% of positive rating. In general, there is more consensus on highly pleasant and unpleasant words than on words in the midrange. The scatterplots for arousal (Fig. 3d) and dominance (Fig. 3e-f) are somewhat similar to that for valence, but less pronounced. Older adults' scatterplot for arousal (Fig. 3c) shows some different patterns, and their responses are distributed in a smaller range than are the younger adults. Finally, the scatterplot of familiarity (Fig. 3g-h) shows that *SD* decreases with increasing means. It shows that there is more consensus on high-familiar words than low-familiar words. Overall, these results are consistent with those of previous studies, in which the perceived valence of

words tends to generalize well whereas the ratings of arousal, dominance and familiarity show greater variability across languages (Eilola & Havelka, 2010; Moors et al., 2013; Soares et al., 2012; Warriner et al., 2013).

Insert Figure 3 here

In order to explore the age variability in ratings, we ran a series of *paired t-test* contrasting average standard deviations between older and younger adults. The arousal (t = 11.11, p < .001, *Cohen's d* = .34) and familiarity (t = 15.47, p < .001, *Cohen's d* = .40) analysis reveals more variability for younger adults than for older adults whereas dominance analysis show more variability for older adults than for younger adults (t = -5.48, p < .001, *Cohen's d* = -.14). For valence, we found similar age-related patterns in average standard deviations (t = .64, p = .53, *Cohen's d* = .01). The scatterplot (Fig. 3a-b) is symmetrical at the median, and this indicates that relative positive or negative words are associated with smaller variability in the ratings across participants, as compared to valence-neutral words (see also Moors et al., 2013; Stadthagen-Gonzalez et al., 2017; Warriner et al., 2013).

3 Reliability of the norms

We explored the interrater reliability of the four ratings with a split-half procedure. First, we randomly split the participants that rated each word into two equal groups and calculated their mean ratings for each word. Second, we computed the Pearson correlations between both sub-groups applying the Spearman–Brown

correction. Third, we repeated these steps 10 times to get a set of 10 correlations. The mean correlation coefficient provided us with the measure of split-half reliability (see Table 3). These steps were repeated for each age group and gender. For all participants, the mean correlations between the two groups are very high for affective dimensions, ranging from a minimum of r = .91 for dominance to a maximum of r = .98 for valence. High correlations are also observed for familiarity, r = .78 (ranging from .77 to .79). The split-half reliabilities for each of the age or gender groups are based on smaller halves than those for all participants, and this may explain why the former are sometimes smaller than the latter. These results show that the ratings are highly reliable and can be used across the entire Chinese speaking population. Regarding these affective variables, valence has a higher interrater reliability than in arousal or dominance ratings, and these findings are in line with previous studies (Ferré et al., 2017; Monnier & Syssau, 2017; Moors et al., 2013; Yao et al., 2017).

Insert Table 3 here

4 Correlations between dimensions

Pearson's correlations, linear and quadratic associations were calculated between dimensions (see Table 4). First, the results show that valence and arousal has the typical U-shaped relationship (see Fig. 4a), which are highly consistent with prior studies (Bradley & Lang, 1999; Eilola & Havelka, 2010; Liu et al., 2018; Schmidtke et al., 2014; Soares et al., 2012; Stadthagen-Gonzalez et al., 2017; Warriner et al.,

2013; Yao et al., 2017). The quadratic relationship between arousal and valence is significant (R^2 = .47, p < .001) and outperforms the linear relationship (R^2 = .033, p < .001). Words that are very positive (e.g., 民富国强 / *Min2fu4guo2qiang2* / *The people are rich and the country is strong*) or very negative (e.g., *丧子之痛* / *Sang4zi3zhi1tong4* / *Bereavement of the son's pain*) are more arousing, as compared to that of valence-neutral words (e.g., 正三角形 / Zheng4san1jiao3xing2 / *Equilateral triangle*). This is corroborated by the positive correlation between valence and arousal for positive words (mean valence rating > 6; r = .40, p < .001) and the negative correlation between them for negative words (mean valence rating < 4; r = -.78, p < .001). Second, Pearson's correlations show that dominance is positively associated with valence (r = .39), but is negatively associated with arousal (r = -.18). The relationships between dominance and valence and between dominance and arousal tend to be linear. However, the linear and quadratic associations do not seem to differentiate much (see Fig. 4b & 4c, Table 4). Words that make people feel happier also make them feel more in control (e.g., 胸怀坦荡 / Xiong1huai2tan3dang4 / Magnanimous mind), and negative words make people feel less in control. Words that make people feel more in control were less arousing (e.g., 实心实意 / Shi2xin1shi2yi4 / Honest and sincere), but words rated less dominant seem to be more arousing (e.g., 天塌地陷 / Tian1ta1di4xian4 / Earth crumbles). Third, Pearson's correlations show that familiarity has positive correlations with valence (r = .23), arousal (r = .054), and dominance (r = .31), although the relationships are nonlinear. As shown in Figure 5, words rated as more familiar are

likely to be regarded as more positive and dominant. Finally, these results should be taken with caution, because they may be mediated by age and gender, which will be considered in detail in the following.

Insert Table 4, Figure 5 here

5 Age-related differences in ratings

To compare emotional ratings across ages, we performed several analyses with mean ratings and RTs as the dependent variables and age as the independent variable. For these 2,061 words, as shown in Table 2, younger adults rate words significantly higher than older adults for dominance (t = 19.26, p < .001, *Cohen's d* = .33) and familiarity (t = 35.00, p < .001, *Cohen's d* = .76), while older adults rate words slightly higher than younger adults for arousal (t = 6.68, p < .001, *Cohen's d* = .12). No age differences are found in mean ratings for valence. Older adults tend to rate words more slowly than do younger adults for all four dimensions (ps < .001). This is in line with previous studies, which indicate that older adults respond more slowly than younger adults (Liu et al., 2015; Paterson et al., 2013; Rayner et al., 2014; Shafto & Tyler, 2014; Stine-Morrow et al., 2006; J. Wang et al., 2018).

The results of gender differences show that females rate words significantly higher than males for arousal (t = 6.10, p < .001, *Cohen's d* =.08), while males rate words slightly higher than females for valence (t = 8.35, p < .001, *Cohen's d* = .05). No gender differences are found in ratings for dominance or familiarity. Specifically,

females tend to rate words more slowly than do males for all four dimensions (*p*s < .001). These results show that younger adults tend to judge words more in control and more familiar than do older adults, and females tend to rate words as more arousing than do males. Since gender differences have been reported much in prior studies (Monnier & Syssau, 2017; Montefinese et al., 2014; Warriner et al., 2013) and sample sizes of each gender were small (appropriately 11 male and 11 female) for each age group, we concentrate on age differences below.

In order to obtain a more detailed picture of our data regarding the impact of age, we grouped the 2,061 words into negative, neutral, and positive words according to the same criteria used in prior studies (Ferré et al., 2012; Warriner et al., 2013; Yao et al., 2017). On the basis of the overall valence score (combined across older and younger adults), we classified words as negative ($M_{valence} \leq 4$), neutral ($4 < M_{valence} \leq 6$), and positive ($M_{valence} > 6$). This procedure resulted in 644 negative, 867 neutral, and 550 positive words. We reported age-related differences in ratings in the following three steps. First, we compared ratings and RTs by older and younger adults. Second, we reported age-related differences in ratings of individual words. Finally, we reported age-related differences in associations between dimensions by older and younger adults.

Age differences in mean evaluations. First, we examined the correlations between older and younger adults' ratings for the 2,061 words. The correlations are extremely high for valence (r = .95), but not for the arousal (r = .73), dominance (r = .62) or familiarity (r = .53) dimensions. It reveals that older and younger adults

agree on whether a word is more positive or more negative than another word. The ratings of arousal, dominance, and familiarity may involve more individual and heterogeneous responses than valence.

Second, in order to further address the question of age differences in emotional ratings, we explored age differences in mean ratings across subsets of negative, neutral and positive words using the following analyses which are introduced by Grühn and Smith (2008). For each rating dimension, we conducted a mixed-design analysis of variance (ANOVA) with age group (older vs. younger) as a within-words factor, and valence category (negative vs. neutral vs. positive) as a between-words factor. Please note that these analyses were performed on the level of words and not on the level of participants. Table 5 shows mean ratings and the results of ANOVAs for older and younger adults across three valence categories. The interaction between age group and valence category, the main effects of age (except valence dimension), and the main effects of valence category are significant for four dimensions (*ps* < .001). Consistent with previous norming studies of older and younger adults (Grühn & Smith, 2008), older adults tend to rate positive words as more positive, more arousing and less controllable than younger adults do. In contrast, older adults tend to rate negative words as more negative, less arousing and more controllable than younger adults do.

Insert Table 5 here

Age differences in mean RTs of evaluations. To compare mean RTs of ratings across age groups, we performed mixed ANOVAs with age (older vs. younger) as a within-words factor, valence category (negative vs. neutral vs. positive) as a between-words factor, word frequency, and word complexity as covariates for each rating dimension. Table 6 and Figure 6 summarize the statistical findings and show the mean RTs for each dimension as a function of age groups and valence category. For the four dimensions, the interaction between age and valence category (except familiarity dimension), the main effects of age, and the main effects of valence category are all significant. Consistent with previous studies of age-related differences in reading (Liu et al., 2015; Stine-Morrow et al., 2006), there are longer RTs of rating for older adults than younger adults partly due to visual and cognitive declines in later life. The interaction between age and valence category was a bit complex (see Fig. 6). For younger adults, negative words tend to be rated more slowly than positive words for the arousal and familiarity dimensions ($p < .001$), and differences between negative and positive words are not significant for valence and dominance dimensions. Neutral words tend to be rated more slowly than positive and negative words for the valence dimension ($p < .001$), whereas the pattern is just the opposite for the arousal dimension ($p < .001$).

For older adults, the patterns are more consistent across four dimensions. Negative words tend to be rated more slowly than positive words for valence, arousal and familiarity dimensions ($ps < .001$). Perhaps the effect of negative stimuli on decision or response stage of word processing is more robust for older adults than

younger adults. Neutral words tend to be rated more slowly than positive words for valence and dominance dimensions, whereas the pattern is just the opposite for arousal and familiarity dimensions (*ps* < .001). Making these abundant age-related affective databases available can help to improve the performance of emotional word recognition models. Currently, emotional factors are conspicuously absent from word recognition models (Citron, et al., 2014; Kuperman et al., 2014).

Insert Table 6 and Figure 6 here

Age differences in ratings of individual words. To address the question of age-related differences in the perception of individual words, we conducted separate independent *t* tests for each word with age as a between-subjects factor. This procedure resulted in 4 (dimensions) × 2,061 (words) = 8,244 analyses on the *t* tests. From the set of 2,061 words, 756 words (36%) show no age-related differences for all four dimensions. However, the other 1,305 words do show difference in at least one dimension. Results show a substantial number of significant main effects of age for valence (413 words; 20% of 2,061 words), arousal (418; 20%), dominance (566; 27%), and familiarity (525; 25%), respectively (see Table 7). These robust age-related differences are mainly from neutral words for valence (197; 48% of 413 words with significant age effects), arousal (212; 51% of 418 words), and familiarity dimensions (230; 44% of 525 words), and are from positive words (246; 43% of 566 words) for dominance dimension. Specifically, for words with significant main effects of age, older adults rate 60% of 197 neutral words as more positive, 77% of 99 positive

words as more arousing, 82% of 92 negative words as more controllable, and 69% of 178 negative words as more familiar than younger adults do. Correspondingly, older adults rate 54% of 134 negative words as more negative, 88% of 107 negative words as less arousing, 99.6% of 246 positive words as less controllable and 94% of 117 positive words as more unfamiliar than younger adults do. These individual words analyses indicate that older adults tend to rate neutral words as more familiar and positive; positive words as more arousing, negative words as more controllable than younger adults do. Correspondingly, older adults tend to rate negative words as less arousing, and positive words as less controllable and unfamiliar than younger adults do. These results are consistent with findings of age-related differences in mean evaluations across subsets of negative, neutral and positive words reported above.

Age differences in associations between dimensions. To examine the relationships between different dimensions and to test whether age influences these relationships, we assessed associations between dimensions for older and younger adults. There are significant age-related differences between the correlation coefficients for valence and arousal (Z = 6.67, p < .001), dominance and arousal (Z = -5.09, p < .001), valence and dominance (Z = -16.16, p < .001), as well as familiarity and valence (Z = 5.90, p < .001). Such difference is not obvious between the correlation coefficients for familiarity and arousal (Z = -1.48, p = .14) as well as familiarity and dominance (Z = 5.90, p = .29). Figure 7 shows the location of each word in a two-dimensional space defined by the mean ratings of each word. These age-related differences observed in word-level data yield the following patterns. First,

compared to younger adults, older adults tend to rate negative words ($M_{valence}$ < 4.27) as less exciting and more in control (see Fig. 7a and 7b). Older adults also tend to rate positive words (4.27 < $M_{valence}$ < 7.73) as more exciting and less dominant. Second, compared to older adults, younger adults have a stronger tendency (r = .55, p < .001) to rate positive words as more in control than negative words (see Fig. 7b). Third, there is a significant negative correlation between dominance and arousal (r_{older} = -.10, $r_{younger}$ = -.25, *ps* < .001). Older adults tend to rate those higher dominant words ($M_{dominance}$ > 3.82) as more exciting, while younger adults tend to rate lower dominant words as more exciting (see Fig. 7c). Forth, older adults show a stronger positive relationship between familiarity and valence than younger adults do (r_{older} = .30, $r_{younger}$ = .12, *ps* < .001). They tend to rate more familiar word ($M_{familiarity}$ > 6.88) as more positive (see Fig. 7d). While pinning down the nature of these age-related differences will be an issue for further investigation, these valuable age-related differences in emotional rating should be considered as potential sources of systematic error or bias for research into emotion words.

Insert Figure 7 here

General discussion

The goal of this study was to establish the CAN4Age norm database and to make these age-related ratings available in the public domain. Although there is a growing body of aging-oriented research on emotion and language, no published word stimulus databases for older adults are available in China. Meanwhile, many studies

frequently use ratings of younger adults to classify stimulus material for older and younger. This would not have taken into account the potential age-related shifts in the perception of material. To address this issue, our work provides valence, arousal, dominance and familiarity ratings of older and younger adults for 2,061 4-character Chinese words. The availability of such a large scale norm database has greatly facilitated the creation of stimulus sets as well as making it possible to include such variables in the automated analysis of text samples (Kratzwald et al., 2018; Liu et al., 2018). With regards to participants' age, the CAN4Age database shows consensus and variation in the perception and meaning of emotional words, which may have implications for current models of word recognition and other applications where the targeted researchers have different profiles.

Associations between dimensions

Consistent with previous research, this work also shows strong associations between dimensions. First, we found the typical U-shaped relationship between valence and arousal, which was reported by many studies (Bradley & Lang, 1999; Eilola & Havelka, 2010; Liu et al., 2018; Schmidtke et al., 2014; Soares et al., 2012; Stadthagen-Gonzalez et al., 2017; Y. Wang et al., 2008; Warriner et al., 2013; Yao et al., 2017). Very positive and very negative words are typically evaluated as highly arousing whereas less emotional and neutral words are less arousing. Second, our results demonstrate that dominance is positively related to valence (Fairfield et al., 2017; Grühn & Smith, 2008; Warriner et al., 2013), indicating that positive words involve a greater degree of control than negative words do. Third, we found

dominance is negative related to arousal, indicating that words rated as less dominant are more arousing (Schmidtke et al., 2014). Forth, familiarity is positively associated with valence (Warriner et al., 2013), arousal, and dominance. It shows that words rated as more familiar are likely to be regarded as more positive, exciting, and strong.

The strength of the correlations between different dimensions may have some implications for the dimensional perspective of emotion since the original model assumes that three dimensions of emotion are orthogonal (Osgood et al., 1957; Russell, 2003; Wundt, 1912/1924). More specifically, the operationalization of dominance may be more complex than has been previously considered. Even though dominance has been identified as an important variable in emotion research, it was much less studied and not often included in previous word norming studies (Gilet et al., 2012; Liu et al., 2018; Monnier & Syssau, 2017; Stadthagen-Gonzalez et al., 2017; Yao et al., 2017). Furthermore, there are inconsistent findings on associations between emotional dimensions. For instance, some studies found that the ratings of dominance and arousal are unrelated (Grühn & Smith, 2008), others found U-shaped relationship (Montefinese et al., 2014; Warriner et al., 2013). In this study, we found dominance and valence are strongly related. This shows that extreme rating values of valence and dominance are more arousing, point again at the utility of considering valence/dominance strength (i.e., how different a word is from neutral) rather than polarity as the explanatory variable (Warriner et al., 2013). It is unclear whether the three rating dimensions could probably be reduced to two latent dimensions (Fontaine, Scherer, Roesch, & Ellsworth, 2007; Grühn & Smith, 2008). Future studies are

needed to validate that dominance explains unique variance over and above valence in information processing. Our database provides the raw data for future study in the dimensional perspective of emotion modeling (e.g., Fontaine et al., 2007; Russell, 2003).

The impact of age: Consensus and variations

With regard to the impact of age, this work shows three major findings. First, different age groups agree on the pleasantness of words, as is evident from the high correlation between older and younger adults' evaluation for valence (although not for the other three dimensions). This indicates that older and younger adults agree on whether a word is more positive or more negative than another word. In line with prior studies (Eilola & Havelka, 2010; Moors et al., 2013; Soares et al., 2012; Warriner et al., 2013), we also found the perceived valence of words tends to generalize well. Yet, the other three dimensions show greater variability across ages. Younger adults rate words significantly higher than older adults for dominance and familiarity. Overall, from the 2,061 words, approximately one third (756 words) of words has no age-related differences in all four dimensions. Two-third, however, shows significant differences.

Second, despite high correlations for valence between older and younger adults, age-related differences are evident for all four rating dimensions. Although the difference in overall mean is generally small between older and younger adults, age-related differences for positive, neutral, and negative words are pronounced.

Older adults tend to evaluate positive words as more positive, more arousing and less controllable than younger adults do. In contrast, older adults tend to rate negative words as more negative, less arousing and more controllable than younger adults do. Older adults tend to give more extreme valence ratings than younger adults do. Older adults' evaluations are more extreme in that positive words are rated more positively and negative words more negatively than that of the younger adults. These findings are consistent with results reported by Grühn and Smith (2008). We also found a stronger relationship between valence and arousal for younger adults than older adults, which is inconsistent with the results reported by Gilet and colleagues (2012) or Ready and colleagues (2017). These findings indicate that the emotional meanings of some words vary with languages. Older and younger adults perceive positive and negative words differently. These age-related differences may be a function of life experience, lifetime exposure, cultural environments, or age-related changes in psychological, biological, and social functioning.

Third, we found complex interactive effects of age and valence category in RTs during rating. Consistent with prior studies (Liu et al., 2015; Paterson et al., 2013; Rayner et al., 2014; Shafto & Tyler, 2014; Stine-Morrow et al., 2006; J. Wang et al., 2018), older adults tend to respond slower than younger adults. Interestingly, we found older adults tend to rate negative words more slowly than positive words for valence, arousal and familiarity dimensions. However, for younger adults, RTs of rating negative and positive words do not significantly differ for valence and dominance dimensions. This pattern of findings may suggest that older and younger

adults process positive and negative words differently. It is possible that the effect of negative stimuli on the decisional or response stage of word processing is more robust for older adults than younger adults. These findings suggest that emotion should be included in models of word recognition as it is likely to make some contribution (Citron, et al., 2014; Kuperman et al., 2014).

Summary, limitations and conclusion

This study provides a large scale emotion norm for 4-character words in Chinese. Analysis to the data demonstrates age-related differences in affective word ratings. While some rating results are consistent with previous studies and across younger and older adults, there are still some differences in ratings for a large number of words. There is a stronger quadratic association between valence and arousal for older adults than younger adults. In general, older adults tend to rate positive words as more positive, more arousing and less controllable, and negative words as more negative, less arousing and more controllable than younger adults do. Overall, older adults tend to give more extreme valence ratings to the words than younger adults do whereas younger adults tend to rate emotional words more controllable and familiar than older adults do.

This work provides an age-adopted tool for future research on the processing of emotion words from a developmental point of view. However, there are some limitations in the present study. First, the materials used do not contain 2-character words, which are ubiquitous in Chinese. Second, this paper purposely did not include

detailed analysis on gender differences because there are small samples of each gender for each age group. Future studies can expand the database to include gender differences. Third, future studies are needed to develop normative databases including discrete emotion ratings for large sets of words which are currently unavailable.

In conclusion, our data set provides a useful source for studies in which the effects of aging are considered and affective words are used. Although some words are evaluated differently by older and younger adults, many words are not. From the acquired set of data, approximately one third (756 words) shows no age-related differences for all four dimensions. Therefore, our collection of affective norms for 2,061 4-character Chinese words gives computational and experimental researchers a much wider selection of materials for their studies. Using the CAN4Age word pool, researchers can select words that are matched across age groups for future affective studies.

Acknowledgements

The authors are grateful to the older and younger adults who participated in the study. This research was partially supported by the National Natural Science Foundation of China (No. 31600887), the Hong Kong Scholars Program (No. XJ2015050), and by the CAS Key Laboratory of Mental Health (No. KLMH2014ZG14). We thank Danlu Liu, Ting Xu, Shuhan Guo, and Juanjuan Yuan for their valuable assistance in recruiting the participants and collecting the data. We thank Xiao Liu, Yike Wang, Yuheng Yang for English translations. We thank Xiaoqing Hu and Shengtao Wu for their comments on an earlier version of this article.

References

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*: (Technical Report C-1). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Carstensen, L. L. (2006). The influence of a sense of time on human development. *Science, 312*(5782), 1913-1915.

Chinese Lexicon. (2003). Beijing, China: [CD-ROM]. Produced by State Key Laboratory of Intelligent Technology and Systems, Tsinghua University and Institute of Automation, Chinese Academy of Sciences. Retrieved from Chinese linguistic Data Consortium (http://www.chineseldc.org/).

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: Time course and ERP correlates. *Neuroscience Letters, 533*, 90-95.

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex Affective Word List. *Applied Psycholinguistics, 35*(2), 313-331.

Eilola, T. M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods, 42*(1), 134-140.

English, T., & Carstensen, L. L. (2014a). Emotional experience in the mornings and the evenings: Consideration of age differences in specific emotions by time of day. *Frontiers in Psychology, 5*, 9.

English, T., & Carstensen, L. L. (2014b). Selective narrowing of social networks across adulthood is associated with improved emotional experience in daily life. *International Journal of Behavioral Development, 38*(2), 195-202.

Fairfield, B., Ambrosini, E., Mammarella, N., & Montefinese, M. (2017). Affective norms for Italian words in older adults: Age differences in ratings of valence, arousal and dominance. *PLoS One, 12*(1), e0169472.

Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. A. (2017). Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. *Behavior Research Methods, 49*(3), 1082-1094.

Ferré, P., Guasch, M., Moldovan, C., & Sánchez-Casas, R. (2012). Affective norms for 380 Spanish words belonging to three different semantic categories. *Behavior Research Methods, 44*(2), 395-403.

Ferrari, V., Bruno, N., Chattat, R., & Codispoti, M. (2017). Evaluative ratings and attention across the life span: Emotional arousal and gender. *Cognition and Emotion, 31*(3), 552-563.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading cognitive state of patients for clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*(12), 1050-1057.

Fung, H. H. (2013). Aging in culture. *Gerontologist, 53*(3), 369-377.

Garrison, K. E., & Schmeichel, B. J. (2019). Effects of emotional content on working memory capacity. *Cognition & Emotion, 33*(2), 370-377.

Gavazzeni, J., Wiens, S., & Fischer, H. (2008). Age effects to negative arousal differ for self-report and electrodermal activity. *Psychophysiology, 45*(1), 148-151.

Gilet, A. L., Grühn, D., Studer, J., & Labouvie-Vief, G. (2012). Valence, arousal, and imagery ratings for 835 French attributes by young, middle-aged, and older adults: The French Emotional Evaluation List (FEEL). *European Review of Applied Psychology, 62*(3), 173-181.

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods, 40*(2), 512-521.

Grühn, D., & Smith, J. (2008). Characteristics for 200 words rated by young and older adults: Age-dependent evaluations of German adjectives (AGE). *Behavior Research Methods, 40*(4), 1088-1097.

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods, 48*, 1358-1369.

Hess, T. M., Popham, L. E., Dennis, P. A., & Emery, L. (2013). Information content moderates positivity and negativity biases in memory. *Psychology and Aging, 28*(3), 853-863.

Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods, 48*(1), 272-284.

Ho, S. M., Mak, C. W., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *PLoS One, 10*(7), e0132294.

Kanske, P., & Kotz, S. A. (2011). Cross-modal validation of the Leipzig Affective Norms for German (LANG). *Behavior Research Methods, 43*(2), 409-413.

Keil, A., & Freund, A. M. (2009). Changes in the sensitivity to appetitive and aversive arousal across adulthood. *Psychology and Aging, 24*(3), 668-680.

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive Affective & Behavioral Neuroscience, 6*(2), 110-126.

Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., & Prendinger, H. (2018). Deep learning for affective computing: Text-based emotion recognition in decision support. *Decision Support Systems, 115*, 24-35.

Kuchinke, L., & Mueller, C. J. (2019). Are there similarities between emotional and familiarity-based processing in visual word recognition? *J Neurolinguistics, 49*, 84-92.

Kunzmann, U., Little, T. D., & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study. *Psychology and Aging, 15*(3), 511-526.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143*(3), 1065-1081.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). *International affective pictures system (IAPS): Digitized photographs, instruction manual and affective ratings (Tech. Rep. A-6)*: Gainesville: University of Florida, NIMH Center for the Study of Emotion and Attention.

Leclerc, C. M., & Kensinger, E. A. (2011). Neural processing of emotional pictures and words: A comparison of young and older adults. *Developmental Neuropsychology, 36*(4), 519-538.

Liu, P., Li, M., Lu, Q., & Han, B. (2018). Norms of valence and arousal for 2,076 Chinese 4-character words. In K. Hasida & W. P. Pa (Eds.), *PACLING 2017* (Vol. 781, pp. 88-98).

Liu, P., Liu, D., Han, B., & Paterson, K. B. (2015). Aging and the optimal viewing position effect in Chinese. *Frontiers in Psychology, 6*, 1-13.

Majerus, S., & D'Argembeau, A. (2011). Verbal short-term memory reflects the organization of long-term memory: Further evidence from short-term memory for emotional words. *Journal of Memory and Language, 64*(2), 181-197.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences, 9*(10), 496-502.

Mathewson, K. J., Arnell, K. M., & Mansfield, C. A. (2008). Capturing and holding attention: The impact of emotional words in rapid serial visual presentation. *Memory & Cognition, 36*(1), 182-200.

Monnier, C., & Syssau, A. (2008). Semantic contribution to verbal short-term memory: Are pleasant words easier to remember than neutral words in serial recall and serial recognition? *Memory & Cognition, 36*(1), 35-42.

Monnier, C., & Syssau, A. (2017). Affective norms for 720 French words rated by children and adolescents (FANchild). *Behavior Research Methods, 49*(5), 1882-1893.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior Research Methods, 46*(3), 887-903.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A.-L., . . . Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods, 45*(1), 169-177.

Mroczek, D. K., & Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology, 75*(5), 1333-1349.

Notthoff, N., & Carstensen, L. L. (2014). Positive messaging promotes walking in older adults. *Psychology and Aging, 29*(2), 329-341.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*: Urbana, IL: University of Illinois Press.

Paterson, K. B., McGowan, V. A., & Jordan, T. R. (2013). Filtered text reveals adult age differences in reading: Evidence from eye movements. *Psychology and Aging, 28*(2), 352-364.

Rayner, K., Yang, J., Schuett, S., & Slattery, T. (2014). The effect of foveal and parafoveal masks on the eye movements of older and younger readers. *Psychology and Aging, 29*(2), 205-212.

Ready, R. E., Santorelli, G. D., & Mather, M. A. (2017). Judgment and classification of emotion terms by older and younger adults. *Aging & Mental Health, 21*(7), 684-692.

Reed, A. E., Chan, L., & Mikels, J. (2014). Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. *Psychology and Aging, 29*(1), 1-15.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145-172.

Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods, 46*(4), 1108-1118.

Shafto, M., & Tyler, L. K. (2014). Language in the aging brain: The network dynamics of cognitive decline and preservation. *Science, 346*(6209), 583-587.

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the affective norms for English words (ANEW) for European Portuguese. *Behavior Research Methods, 44*(1), 256-269.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.

Stadthagen-González, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598-605.

Stadthagen-Gonzalez, H., Imbault, C., Sánchez, M. A. P., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods, 49*, 111-123.

Steenhaut, P., Demeyer, I., De Raedt, R., & Rossi, G. (2018). The role of personality in the assessment of subjective and physiological emotional reactivity: A comparison between younger and older adults. *Assessment, 25*(3), 285-301.

Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods, 39*(4), 1020-1024.

Stine-Morrow, E. A. L., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin, 132*(4), 582-606.

Streubel, B., & Kunzmann, U. (2011). Age differences in emotional reactions: Arousal and age-relevance count. *Psychology and Aging, 26*(4), 966-978.

Uchino, B. N., Birmingham, W., & Berg, C. A. (2010). Are older adults less or more physiologically reactive? A meta-analysis of age-related differences in cardiovascular reactivity to laboratory tasks. *Journals of Gerontology: Psychological Sciences, 65B*(2), 154-162.

Wang, J., Li, L., Li, S., Xie, F., Chang, M., Paterson, K. B., . . . McGowan, V. A. (2018). Adult age differences in eye movements during reading: The evidence from Chinese. *Journals of Gerontology: Psychological Sciences, 73*(4), 584-593.

Wang, Y., Zhou, L. M., & Luo, Y. J. (2008). The pilot establishment and evaluation of Chinese Affective Word System. *Chinese Mental Health Journal, 22*, 39-43.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191-1207.

Wechsler, D. (1981). *WAIS-R Manual: Wechsler Adult Intelligence Scale-revised*. New York: Psychological Corporation.

Wieser, M. J., Muhlberger, A., Kenntner-Mabiala, R., & Pauli, P. (2006). Is emotion processing affected by advancing age? An event-related brain potential study. *Brain Research, 1096*, 138-147.

Wirth, M., Isaacowitz, D. M., & Kunzmann, U. (2017). Visual attention and emotional reactions to negative stimuli: The role of age and cognitive reappraisal. *Psychology and Aging, 32*(6), 543-556.

Wolf, T., & Demiray, B. (2019). The mood-enhancement function of autobiographical memories: Comparisons with other functions in terms of emotional valence. *Consciousness and Cognition, 70*, 88-100.

Wundt, W. M. (1912/1924). *An introduction to psychology (R. Pintner, Trans.)*: London: Allen & Unwin. (Original work published 1912).

Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods, 49*(4), 1374-1385.

Appendix: Instructions

The original Chinese instructions are presented as well as their English translation.

欢迎您参与本研究！请根据您的第一感觉和经验来评价词语带给您的感受。请您在9点量表上对同一个词进行两步评价，用鼠标左键准确点击最能代表您直观感受的相应数字。选择没有正确错误之分，没有时间限制，但是请您不要花费太多时间考虑这些词。

You are invited to take part in the study that is investigating how people respond to different types of words. Your will use a 9-point scale to rate how you felt while reading each word in two steps. There were no right and wrong answers, and the best answer would reflect your true opinion about the word. Please make your ratings based on your first and immediate reaction by clicking on the appropriate figure using a computer mouse. Please work at your own pace and don't spend too much time thinking about each word.

Instructions for rating valence and arousal (Adapted from Bradley & Lang, 1999; Stadthagen-Gonzáez et al., 2017; Wang et al., 2008; Warriner et al. 2013)

第一步：请您评价看到词语后的愉快程度，点击一下相应数字。1表示极其不悦、极其烦恼、极度不满意、很忧伤或非常失望等负性感受。9表示极其高兴、极其愉快、极其满意或充满希望等正性感受或正能量。其中，2=非常不愉悦，3=比较不愉悦，4=有点不愉悦，5=感觉一般，没有任何负性或者正性的感受，6=有点愉悦，7=比较愉悦，8=非常愉悦。

第二步：请评价您看到同一个词后的心情激动程度，点击一下相应数字。1表示看词后感到极其平静放松、不警觉、极少刺激性、引起的关注量最少等。9表示看词后感到极其激动、够刺激、令人非常觉醒、极其兴奋、眼前一亮或者心情起伏波动非常大等。其中，2=非常不平静，3=比较不平静，4=有点不平静，5=感觉一般，6=有点激动，7=比较激动，8=非常激动。

In the first step, please judge the extent to which the words referred to something that is positive/pleasant or negative/unpleasant using a 9-point scale. At one extreme of the scale, you are completely unhappy, annoyed, unsatisfied, melancholic, despaired or bored. When a word makes you feel extremely unhappy, you should indicate it by selecting 1. The other end of the scale is for when you feel extremely happy, pleased, satisfied, contented, or hopeful. When a word makes you feel extremely happy, you should indicate it by selecting 9. The other numbers on the scale also allow you to describe your intermediate feelings of pleasure when you read each word (2=very negative/unpleasant, 3=moderately negative/unpleasant, 4=slightly negative/unpleasant, 5=neither happy nor sad, 6=slightly positive/pleasant, 7=moderately positive/pleasant, 8= very positive/pleasant).

In the second step, please judge the extent to which the words referred to something that is calm or excited using a 9-point scale. At one extreme of the scale, you are completely relaxed, calm, sluggish, dull, or sleepy. When a word makes you feel totally calm, you should indicate it by selecting 1. The other end of the scale is for when you feel stimulated, excited, frenzied, jittery, wide-awake, or aroused. When a word makes you feel totally excited, you should indicate it by selecting 9. The other

numbers on the scale also allow you to describe intermediate feelings of calmness/arousal (2=very calm, 3=moderately calm, 4=slightly calm, 5=neither calm nor excited, 6=slightly excited, 7=moderately excited, 8=very excited).

Instructions for rating dominance and familiarity (Adapted from Bradley & Lang, 1999; Wang et al., 2008; Warriner et al. 2013)

第一步：请您评价看到词语后的控制支配感。从1到9表示控制感越来越强。1表示感到自己完全地受控制、受影响、感到敬畏、感到弱小顺从、被引导、受约束、受操控的、屈服的、无能为力的。9表示感到自己完全处于支配地位、具有充分的支配权和控制感，完全能制约的、能操纵的、有影响力的、有重要地位的、居优势地位的、有自主权的。5表示一般，即您感到既不受控制，也没有感受到支配地位，您可以选择“5”。其中，2=非常受支配，3=比较受支配，4=有点受支配，5=感觉一般，6=有点控制感，7=比较控制感，8=非常控制感。

第二步；请您评价对于这个词汇的熟悉程度，从1到9熟悉度越来越强。1表示阅读词汇后感到强烈的陌生和生疏感、不认识这个词，从来没有见过、看过或者用过这个词。9表示对这个词极其的熟悉、熟知、亲切、经常使用等，比如通过语言交流、网络、电视和书籍等多种方式见过。其中，2=非常不熟悉，3=比较不熟悉，4=有点不熟悉，5=感觉一般，6=有点熟悉，7=比较熟悉，8=非常熟悉。比如看到词汇“自贻伊戚”，您可能从来没有见过，就可以选择数字“1”；而看到词汇“美味米饭”，您可能每天都见到。

In the first step, please judge the extent to which the words referred to something that is weak/submissive or strong/dominant using a 9-point scale. At one

extreme of the scale, you are completely influenced, cared-for, awed, submissive, or guided. When a word makes you feel completely submissive, you should indicate it by selecting 1. The other end of the scale is for when you feel extremely in control, influential, important, dominant, autonomous, or controlling. When a word makes you feel completely dominant, you should indicate it by selecting 9. The other numbers on the scale also allow you to describe your intermediate feelings of control when you read each word (2=very weak/submissive, 3=moderately weak/submissive, 4=slightly weak/submissive, 5=neither submissive nor dominant, 6=slightly strong/dominant, 7= moderately strong/dominant, 8= very strong/dominant).

In the second step, please judge the familiarity which involves rating how often the given word occurs in everyday language in either written or spoken form using a 9-point scale. At one extreme of the scale, you are completely familiar with this word in everyday language. Maybe you often hear the word on conversation, at the radio, at TV, or you may find it in a written form books, Internet, etc. When a word makes you feel totally familiar, you should indicate it by selecting 9. Conversely, a sore of 1 indicates that you rarely find the word in everyday language. The other numbers on the scale also allow you to describe intermediate feelings of familiarity (2=very unfamiliar, 3=moderately unfamiliar, 4=slightly unfamiliar, 5=undecided, 6=slightly familiar, 7= moderately familiar, 8= very familiar). For example, the word 美味米饭 (“delicious rice”) could be rated as occurring in everyday language very often, whereas the word 自贻伊戚 (“torture oneself”) could be rated as never occurring in everyday language.

最后再举例：比如看到词汇“国家军队”、“打架斗殴”、“蓝天白云”、“桌子椅子”等，请按照您的直觉体验感受，在相应维度上选择相应的数字。

Finally, we provide examples (e.g., national army, fight and hit, blue sky and white clouds, tables and chairs) for the ratings of relevant dimensions. According to your first and immediate reaction, which figures would you select?

Table 1. Demographic characteristics and neuropsychological performance of the older and younger participants (mean and standard deviations)

	Older (n = 114)	Younger (n = 150)	*p* [a]
Age	70.05 (6.01)	21.59 (3.41)	<.001
Gender (Female/Male)	62/52	75/75	-
Education (in years)	13.61 (3.06)	14.30 (2.25)	.037
Self-rated health [b]	4.77 (1.19)	5.32 (1.15)	<.001
MMSE [c]	29.24 (.91)	29.69 (.64)	<.001
Digit span forward	7.81 (1.02)	9.85 (1.07)	<.001
Digit span backward	5.25 (.90)	7.88 (1.27)	<.001
Vocabulary test	57.61 (6.78)	57.55 (8.27)	.95
Verbal fluency test	19.74 (4.93)	23.96 (5.33)	<.001

[a] *p*: Independent samples two-tailed *t*-tests

[b] Self-rated health as measured on a 7-point scale: 1 = very poor; 2 = fairly poor; 3 = somewhat poor; 4 = neutral; 5 = somewhat good; 6 = fairly good; 7 = very good

[c] *MMSE* refers to the Mini Mental State Examination

Table 2. Descriptive statistics and group differences for valence, arousal, dominance, and familiarity ratings and RTs by age and gender

	Mean	*AvgSD*	Min	Max	Range	Mean	*AvgSD*	Min	Max	Range	*p*
	Older_Ratings					**Younger_Ratings**					
Valence	4.89	1.10	1.35	8.07	6.72	4.88	1.11	1.79	7.95	6.16	.38
Arousal	5.63	1.37	4.29	8.00	3.71	5.54	1.50	3.00	7.83	4.83	<.001
Dominance	4.72	1.74	2.17	6.83	4.66	5.07	1.69	2.05	7.68	5.63	<.001
Familiarity	6.84	1.22	4.78	8.04	3.26	7.20	1.37	4.44	8.42	3.98	<.001
	Female_Ratings					**Male_Ratings**					
Valence	4.86	1.15	1.38	7.92	6.54	4.92	1.09	1.72	7.69	5.97	<.001
Arousal	5.61	1.48	3.19	8.38	5.19	5.56	1.46	3.83	7.65	3.82	<.001
Dominance	4.90	1.77	2.09	7.27	5.18	4.89	1.76	1.79	7.25	5.46	.51
Familiarity	7.02	1.40	4.48	8.12	3.64	7.02	1.23	4.83	8.13	3.30	.49
	Older_RTs (s)					**Younger_RTs (s)**					
Valence	5.40	2.89	1.71	9.10	7.39	2.88	1.63	1.56	5.46	3.90	<.001
Arousal	2.50	2.16	.44	6.57	6.13	1.25	1.00	.47	2.68	2.22	<.001
Dominance	6.17	3.73	3.40	10.0	6.60	3.52	2.54	1.82	6.53	4.70	<.001
Familiarity	1.72	1.26	.97	3.47	2.50	1.07	.99	.54	2.14	1.60	<.001
	Female_RTs (s)					**Male_RTs (s)**					
Valence	4.21	2.78	1.76	8.01	6.25	4.09	2.55	1.77	7.19	5.42	<.001
Arousal	2.04	1.93	.50	4.93	4.43	1.72	1.61	.41	3.94	3.53	<.001
Dominance	4.87	3.59	2.88	8.16	5.27	4.81	3.36	2.62	7.46	4.85	<.001
Familiarity	1.42	1.23	.75	2.66	1.91	1.36	1.14	.69	2.85	2.16	<.001

Reported are the group means (i.e., ratings and RTs), the average standard deviations (*AvgSD*), the minimum (Min), the maximum (Max), and the range of the average rating means, and, in the last column, the *p* value of a two-tailed paired *t* test comparing the group means.

Table 3. Means (M) and range for the interrater split-half reliabilities for each dimension by age and gender

	All participants		Older		Younger		Female		Male	
Dimension	M	Range	M	Range	M	Range	M	Range	M	Range
Valence	.99	.99~.99	.98	.98~.98	.98	.97~.98	.98	.97~.98	.98	.97~.98
Arousal	.92	.91~.92	.80	.79~.81	.89	.88~.90	.86	.85~.86	.82	.81~.83
Dominance	.91	.90~.91	.80	.80~.82	.88	.88~.88	.83	.82~.85	.83	.82~.88
Familiarity	.78	.77~.79	.62	.61~.65	.70	.69~.73	.61	.59~.63	.69	.68~.70

Table 4. Pearson correlations (r), linear and quadratic associations between dimensions for all, older and younger adults

	r	Linear			Quadratic			
		R^2	*F*	*b*	R^2	*F*	b_1	b_2
All								
Val vs. Aro	-.18***	.033	69.24	-.089	.47	908.8	-2.40	.24
Dom vs. Aro	-.18***	.032	67.12	-.15	.038	40.98	-.77	.064
Val vs. Dom	.39***	.15	361.12	.22	.15	181.32	.33	-.011
Fam vs. Val	.23***	.051	111.68	.77	.089	101.05	-12.81	1.00
Fam vs. Aro	.054*	.0029	5.96	.09	.013	13.03	3.45	-.25
Fam vs. Dom	.31**	.098	224.41	.62	.13	157.16	-6.87	.55
Val vs. Aro								
Older	-.048*	.0023	4.80	-.02	.53	1159	-2.01	.21
Younger	-.25***	.062	136.02	-.16	.38	624.34	-2.73	.27
Dom vs. Aro								
Older	-.10***	.010	21.03	-.075	.014	14.56	-.50	.046
Younger	-.25***	.064	141.34	-.22	.073	81.20	-.84	.062
Val vs Dom								
Older	.11***	.012	24.07	.06	.013	13.27	.18	.013
Younger	.55***	.301	886.76	.40	.301	443.85	.49	-.010
Fam vs. Val								
Older	.30***	.089	201.99	1.06	.12	142.88	-11.18	.91
Younger	.12***	.015	31.39	.32	.032	33.89	-4.89	.38
Fam vs. Aro								
Older	.075**	.0056	11.61	.11	.0058	6.01	.51	-.030
Younger	.061**	.004	7.81	.10	.015	15.50	2.76	-.19
Fam vs. Dom								
Older	.26***	.065	142.64	.50	.071	78.47	-2.43	.22
Younger	.23**	.053	114.24	.44	.088	99.31	-5.03	.40

Val = Valence, Aro = Arousal, Dom = Dominance, Fam = familiarity

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 5. Older and younger adults' mean ratings (SD) and results of ANOVAs for 644 negative, 867 neutral and 550 positive words

	Negative		Neutral		Positive		Age × VC		Age		VC	
	Older	Younger	Older	Younger	Older	Younger	*F*	η^2	*F*	η^2	*F*	η^2
Valence	3.10(.59)	3.19(.51)	5.11(.57)	5.04(.61)	6.65(.43)	6.61(.42)	23.75***	.023	.15	.00	7734***	.70
Arousal	5.85(.60)	6.09(.65)	5.26(.43)	5.04(.82)	5.95(.55)	5.69(.80)	150.80***	.13	42.22***	.02	410.74***	.29
Dominance	4.66(.77)	4.39(.86)	4.67(.76)	5.12(.91)	4.87(.91)	5.78(.83)	456.78***	.31	562.08***	.22	161.30***	.14
Familiarity	6.73(.36)	7.18(.47)	6.78(.43)	7.12(.60)	7.05(.37)	7.37(.47)	13.43***	.013	1204.64***	.37	81.09***	.073

F and η^2 values are for the interaction between age group and valence category (age × VC), as well as for the main effects of age group and valence category (VC). Please note that 2,061 words were classified on the basis of the overall valence score. Thus, some words might be classified in different valence groups when older and younger adults' ratings are considered, respectively. *** $p < .001$

Table 6. Statistical values for ANOVAs analyses between older and younger adults' mean response times for 644 negative, 867 neutral and 550 positive words

	Age × VC		Age		VC		Age × WordFre		Age × WordCom		WordFre		WordCom	
	F	η^2	*F*	η^2	*F*	η^2	*F*	η^2	*F*	η^2	*F*	η^2	*F*	η^2
Valence	13.93***	.013	535.74***	.21	39.85***	.037	3.23+	.002	11.48**	.006	9.28**	.004	8.19**	.004
Arousal	470.13***	.317	308.40***	.13	677.06***	.40	1.69	.19	19.09***	.009	4.09*	.002	17.74***	.009
Dominance	18.17***	.017	672.82***	.25	14.54***	.014	.58	.00	1.21	.00	7.12**	.003	.02	.00
Familiarity	1.45	.001	395.56***	.16	25.34***	.024	3.05+	.001	2.66	.001	6.37*	.003	3.36+	.002

F and η^2 values are for the interaction between age group and valence category (age × VC), as well as for the main effects of age group and valence category (VC). Since word frequency (WordFre) and word complexity (WordCom) were regarded as covariates for ANOVAs analyses, the interactions between age and word frequency (Age × WordFre) and between age and word complexity (Age × WordCom), the main effects of word frequency (WordFre) and word complexity (WordCom) were also reported. *** $p < .001$, ** $p < .01$, * $p < .05$, + $.05 < p < .10$

Table 7. The number of prominent words (percentage) with significant age differences on four dimensions by valence category (negative, neutral, and positive), respectively

	Negative	Neutral	Positive	Total
Valence	134 (32%)	**197 (48%)**	82 (20%)	413 (100%)
Arousal	107 (25%)	**212 (51%)**	99 (24%)	418 (100%)
Dominance	92 (16%)	228 (40%)	**246 (44%)**	566 (100%)
Familiarity	178 (34%)	**230 (44%)**	117 (22%)	525 (100%)

Figure 1

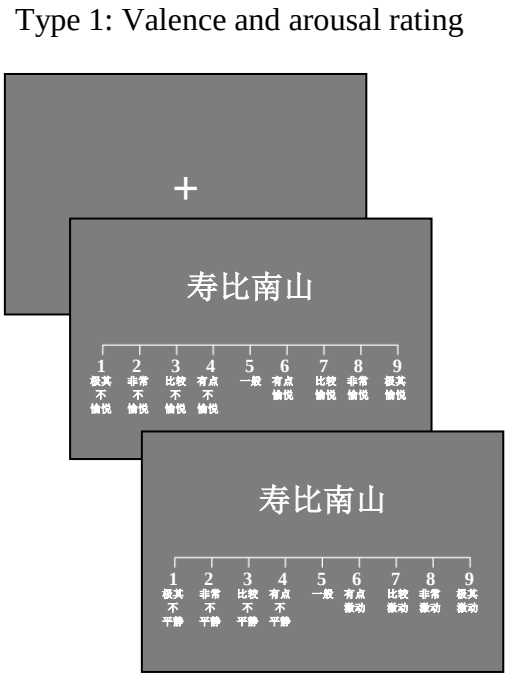


Fig. 1 An example of the paradigm used in the study to explore the ratings of relevant dimensions for the 4-character word 寿 比 南 山 (Shou4bi3nan2shan1/longevity). Each trial began with a fixation cross (+) displayed in the center of the screen for 600 ms. Then the given word and the respective 9-point scale were presented on the screen until participants responded by clicking on the appropriate figure to make their rating using the computer mouse. In data collection of Type 1 questionnaire, participants rated all of the words first for valence, and then for arousal. In data collection of Type 2 questionnaire, participants rated first for dominance, and then for familiarity. Response scales ranged from extremely unpleasant (1) to extremely pleasant (9) for valence, from extremely calming (1) to extremely exciting (9) for arousal, from extremely controlled (1) to extremely control (9) for dominance, and from extremely unfamiliar (1) to highly familiar (9). These instructions in detail are provided in the Appendix.

Figure 2

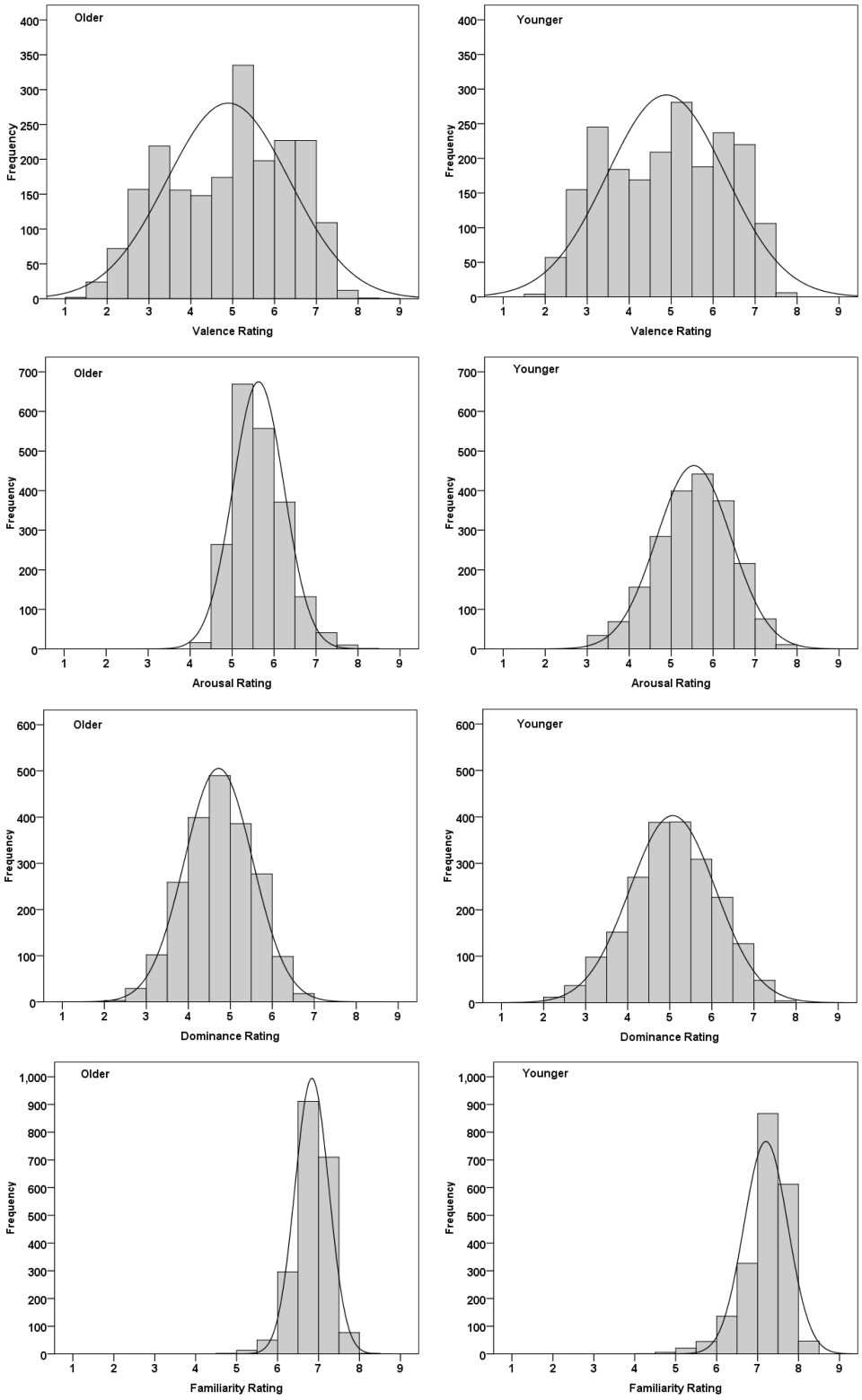


Fig.2 Distribution of valence, arousal, dominance and familiarity ratings for the older and younger adults. Each bar represents the number of words rated within one interval of the scale. The theoretical normal curves are shown by solid lines.

Figure 3

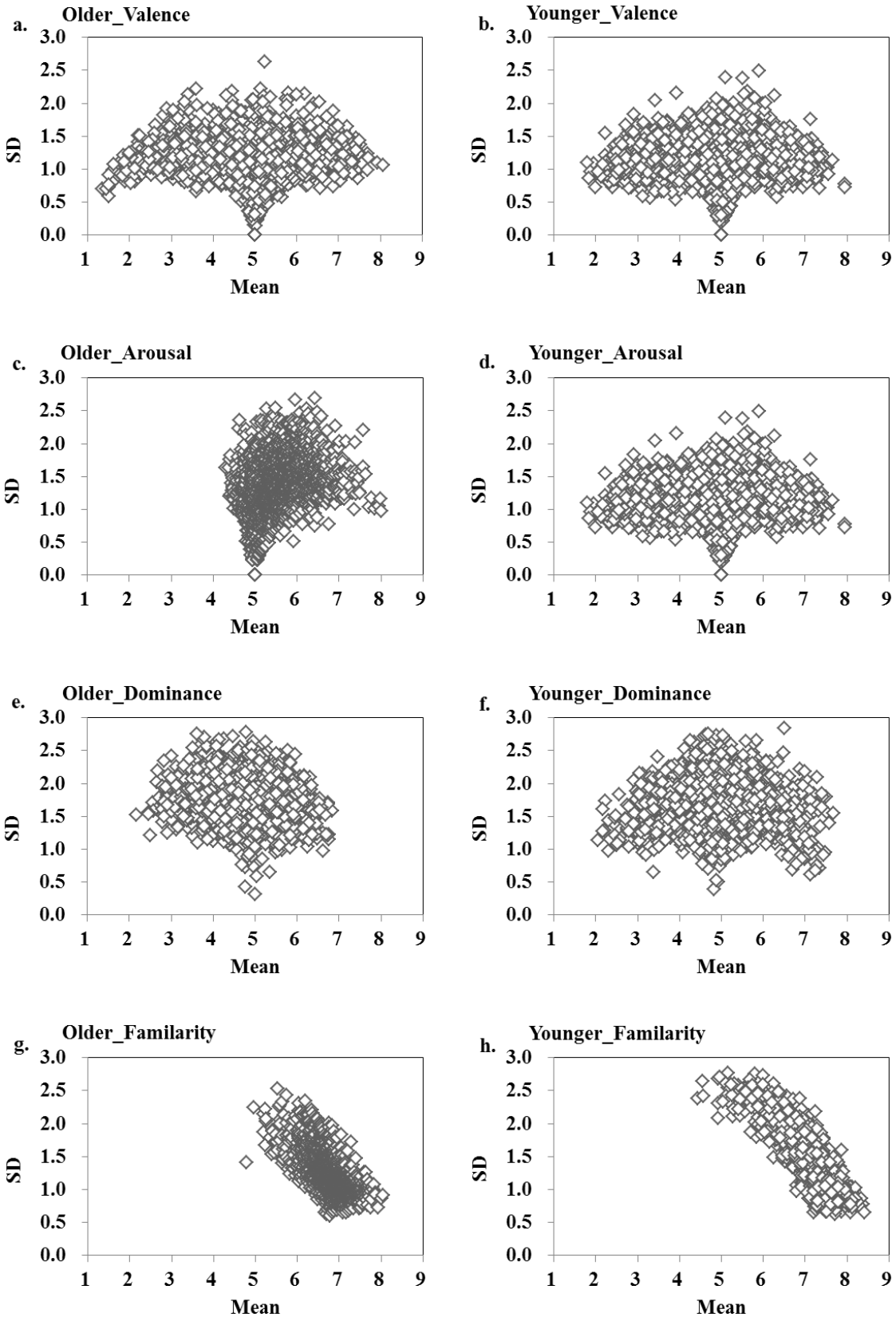
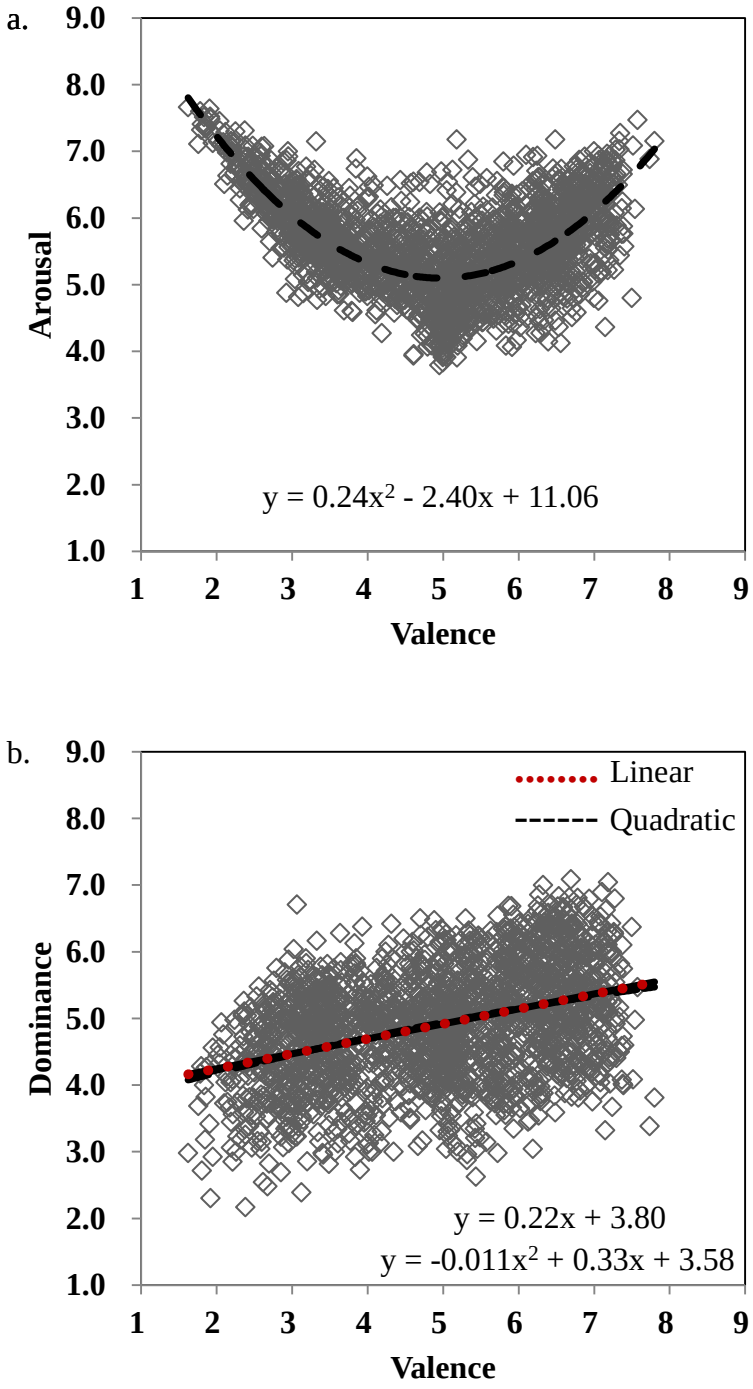


Fig.3 Average standard deviations (variance among responders) across the valence, arousal, dominance and familiarity ranges per word for older and younger adults

Figure 4



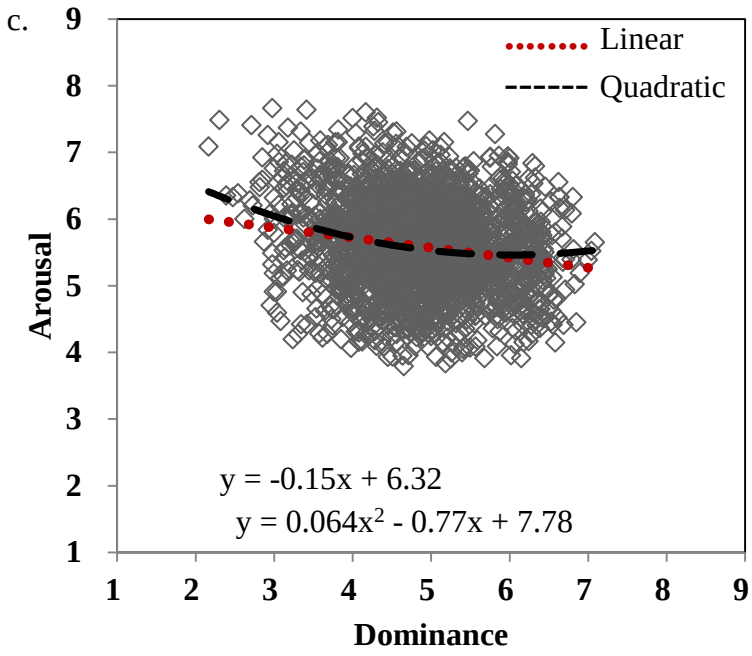


Fig.4 Scatterplots of dimensions (*a*, arousal vs. valence; *b*, dominance vs. valence; *c*, Arousal vs. dominance) for all 2,061 words. The linear and quadratic associations between dimensions are represented by red dotted and black dashed lines, respectively.

Figure 5

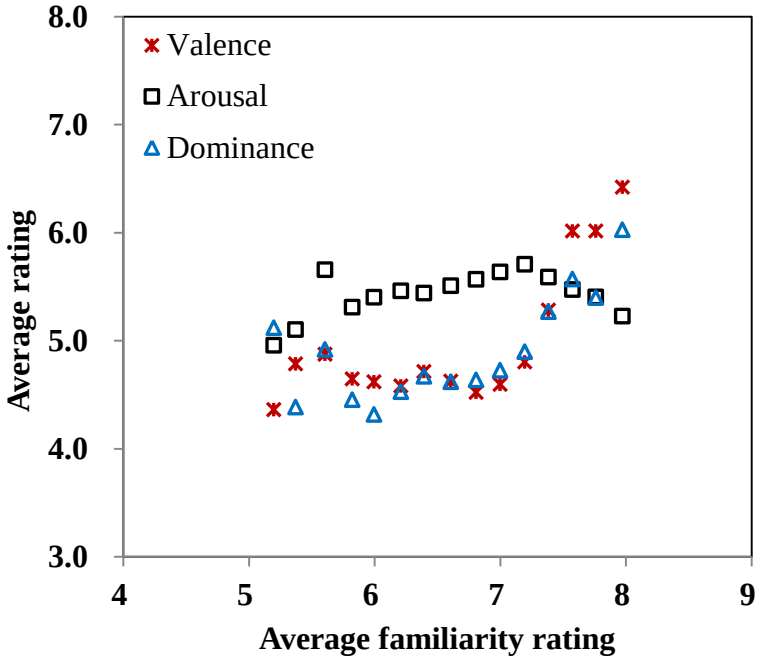
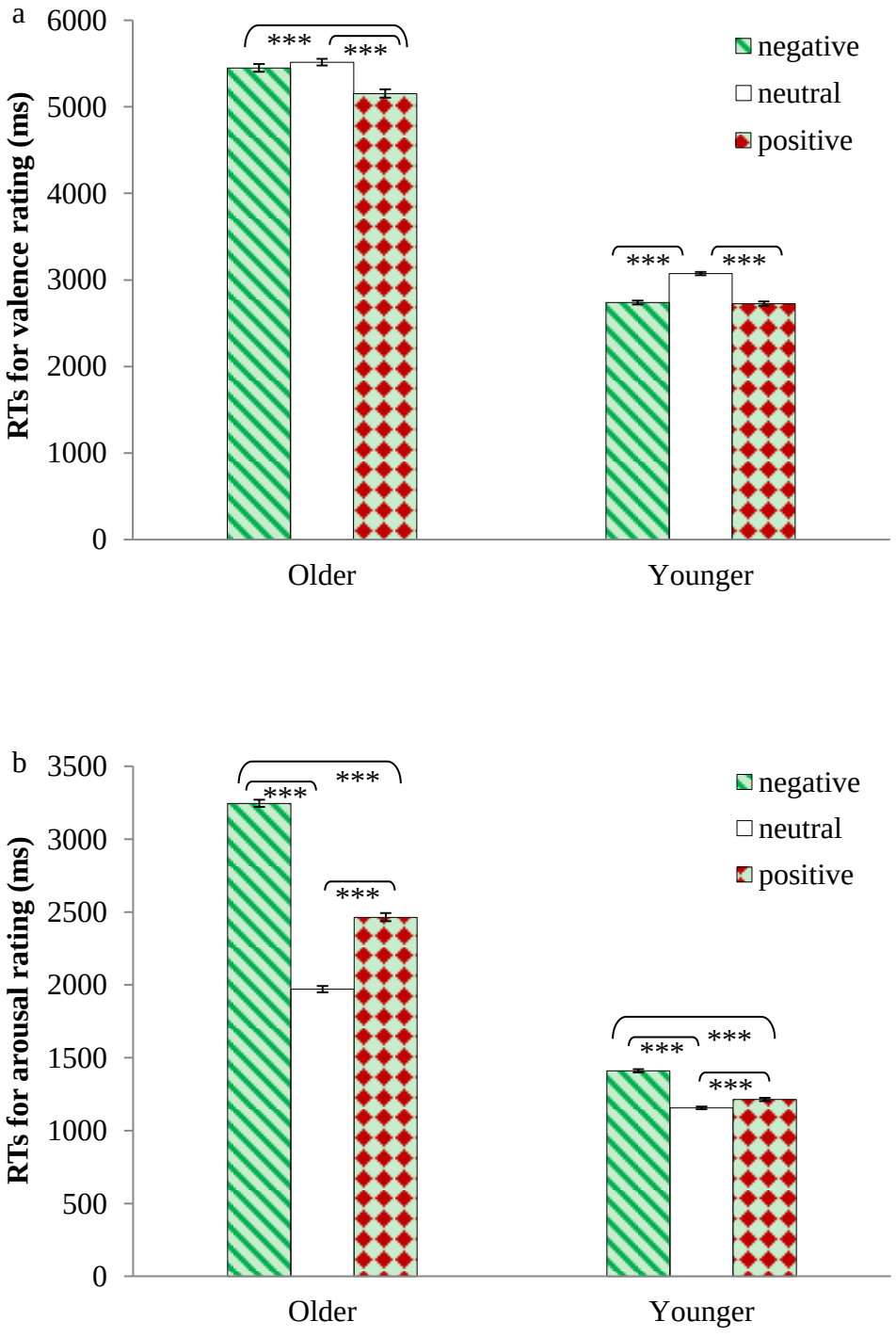


Fig.5 Relationships between three affective dimensions and familairity for overal average rating.

Figure 6



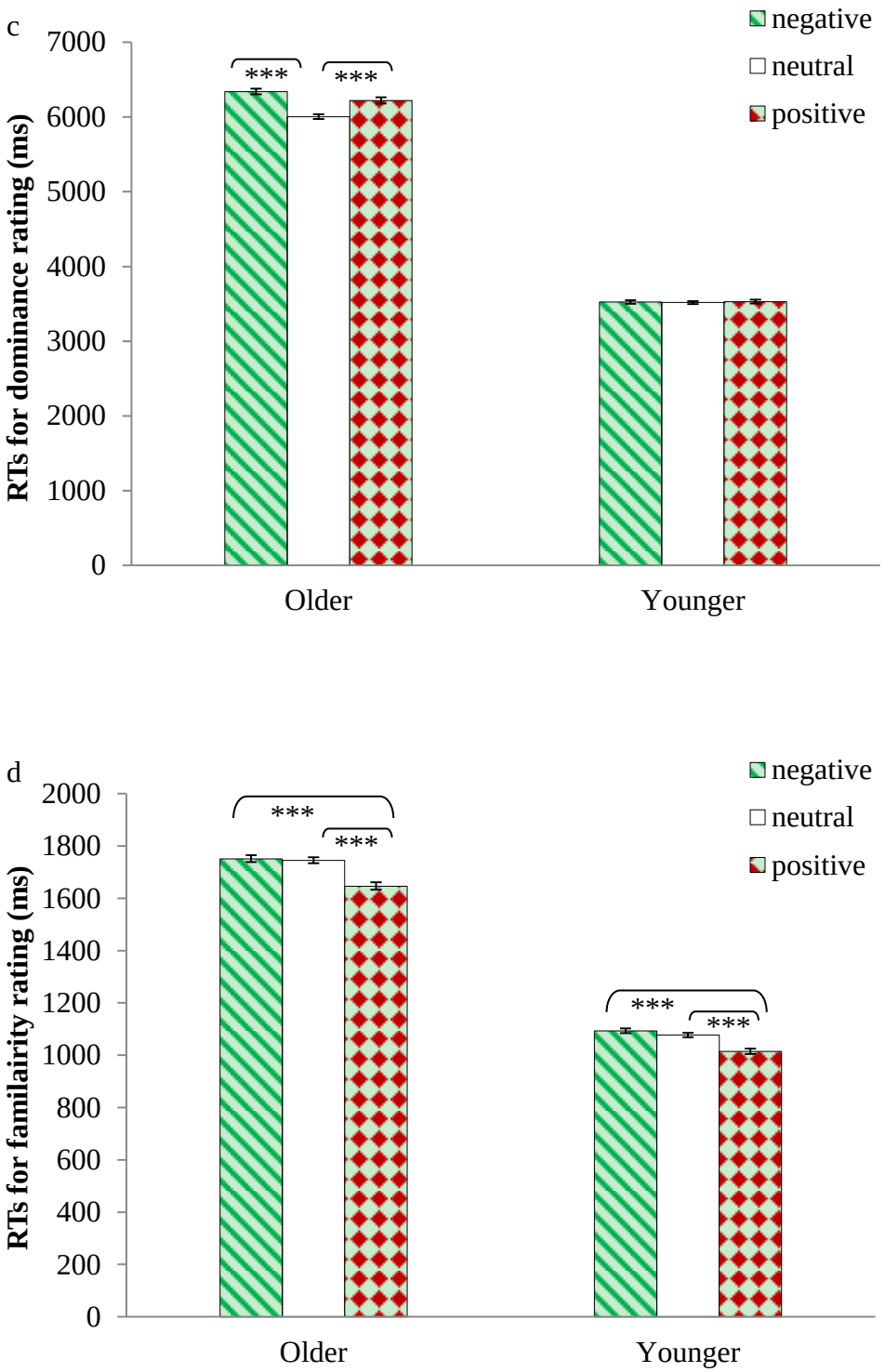
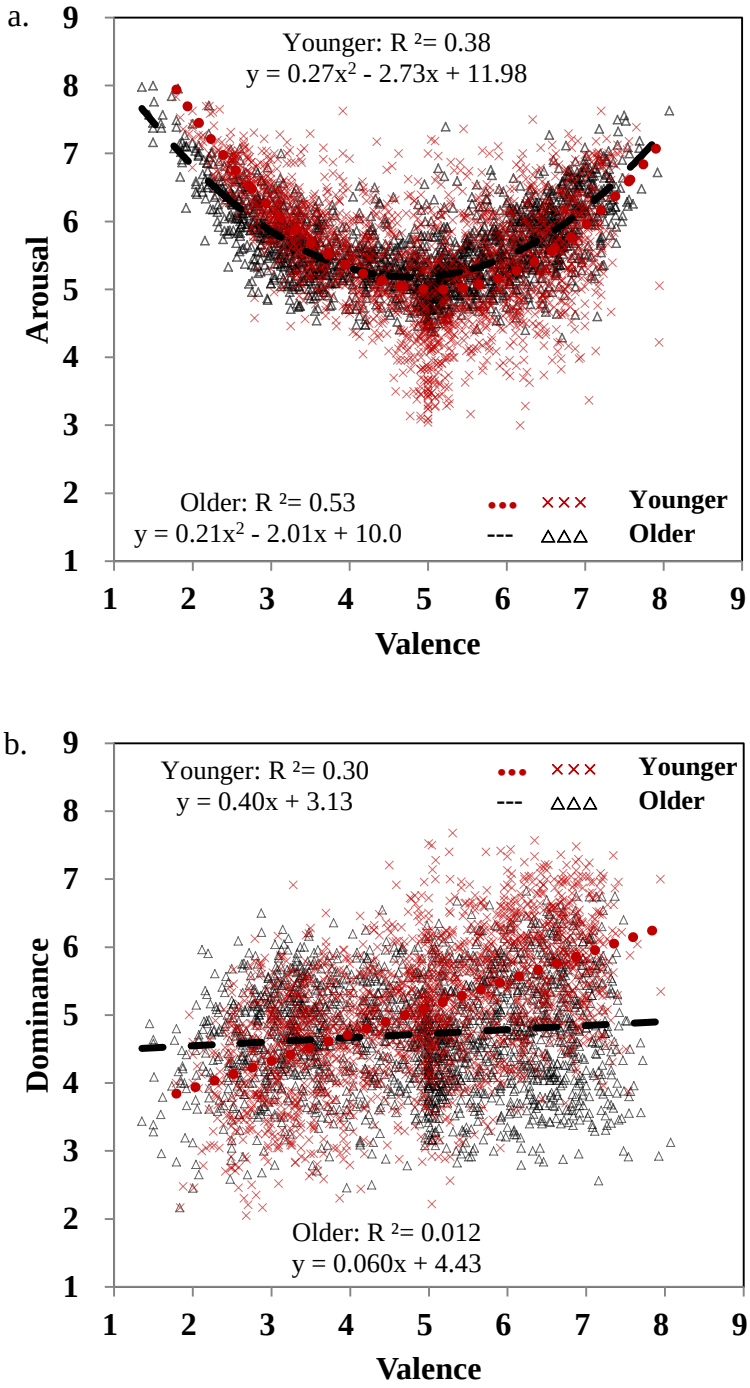


Fig.6 Means and standard errors for RTs of ratings in the four emotional dimensions (a, valence; b, arousal; c, dominance; d, familiarity) for older and younger adults as a function of valence category (negative, neutral, positive). *** $p < .001$

Figure 7



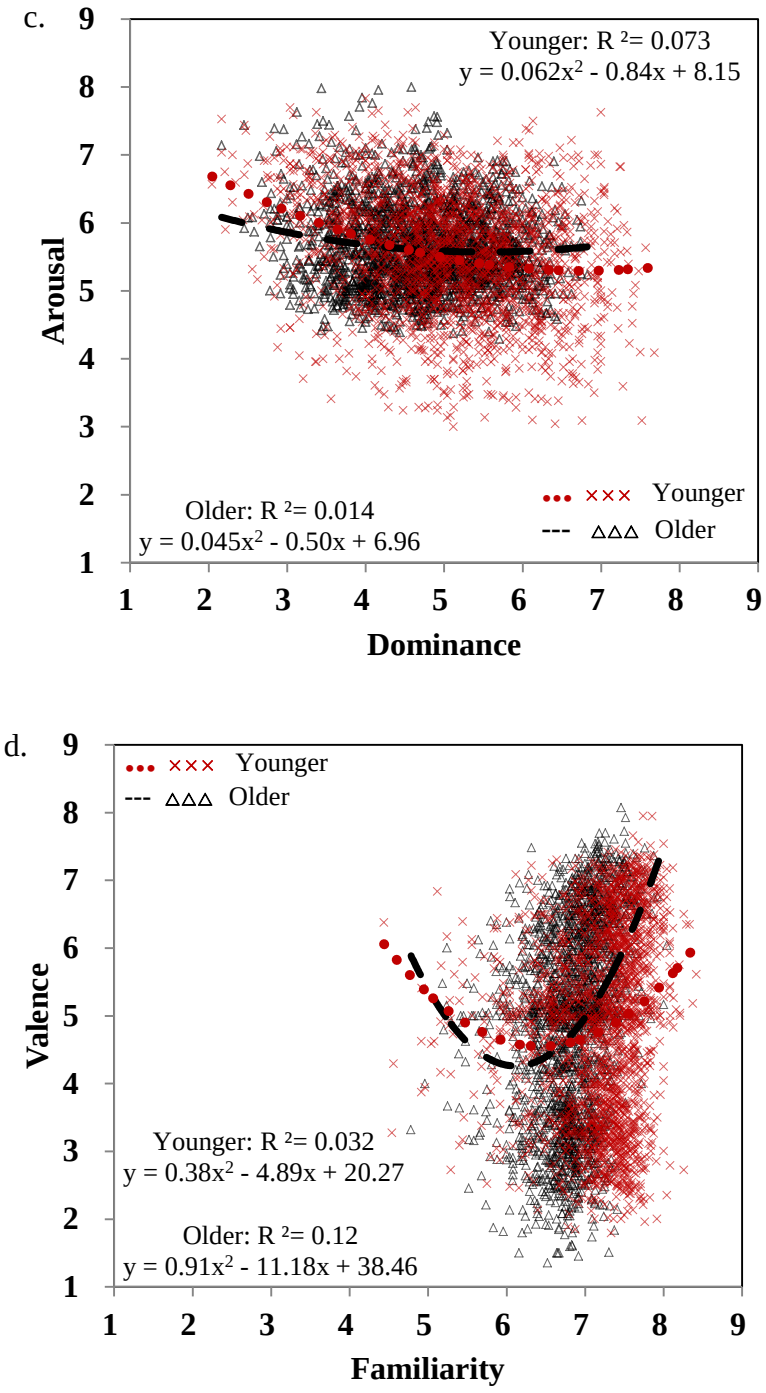


Fig. 7 Age differences in the scatterplots of dimensions (*a*, valence vs. arousal; *b*, valence vs. dominance; *c*, dominance vs. arousal; *d*, familiarity vs. valence) for all 2,061 words (△ in black for older, × in red for younger). The best fit of regressions lines between dimensions are shown by black dashed and red dotted lines for older and younger adults, respectively.